

# COVID-19 no Brasil: uma abordagem via mineração de texto no *Twitter*

Gabriel Baldasso <sup>1</sup>

Anderson C. Soares de Oliveira <sup>2</sup>

Eric Batista Ferreira <sup>3</sup>

## Resumo

A pandemia da COVID-19 que teve início no fim de 2019, provocou o maior isolamento social já antes visto, com o intuito de frear a propagação do vírus, o que trouxe um aumento de usuários da internet e das redes sociais. Somente no Brasil, o *Twitter* apresenta um crescimento de 28% entre seus usuários de 2020 para 2021, o que acarreta um aumento no número de publicações a cada segundo. A interação dos usuários por meio da internet deixa rastros digitais sobre o seu cotidiano e comportamento, o que levou esses ambientes virtuais a serem objetos de estudo de diversas áreas do conhecimento. Contudo, como estamos diante de um *BigData*, técnicas de mineração de texto são essenciais para extrair dados da rede. Logo, neste trabalho tem-se o objetivo de monitorar o uso das palavras-chave: "covid", "vacina", "pandemia" e "isolamento" entre os usuários brasileiros do *Twitter*, e compará-las com o número de casos, de mortes e de vacinados com a primeira dose fornecidos pelo Ministério da Saúde. Para isso, uma amostra de 5 milhões de *tweets* foi coletada de 3000 contas ativas do *Twitter*, por meio do software R, com o pacote *rtweet*. As palavras-chave foram filtradas para o período de janeiro de 2020 a junho de 2021. No conjunto de *tweets* coletados: 1,00% citavam a palavra "covid"; 0,96% "vacina", 0,56% "pandemia", e 0,10% "isolamento". O número de casos, mortes e vacinados com a primeira dose, apresentaram correlação positiva acima de (0,80) com os termos "covid" e "pandemia", com a maior correlação (0,9257) estando entre número de mortes e o termo "pandemia". O menor valor de correlação que se destacou foi o termo "isolamento" que apresentou correlação de (0,3956) com o número de vacinados. Dessa forma, pode-se concluir que o uso dos termos relacionados a pandemia da COVID-19 no *Twitter*, são correlacionados positivamente aos dados provenientes do Ministério da Saúde.

**Palavras-chave:** Covid; Pandemia; Vacina; Twitter; Mineração de texto.

<sup>1</sup>Mestrando em Estatística Aplicada e Biometria (UNIFAL-MG) - gabrielbaldasso@gmail.com

<sup>2</sup>Professor no Instituto de Ciências Exatas e da Terra (UFMT) - andersoncso@gmail.com

<sup>3</sup>Professor do Departamento de Estatística da Universidade Federal de Alfenas (Unifal-MG) - eric.ferreira@unifal-mg.edu.br